



September 20, 2019

**VIA ELECTRONIC FILING**

Marlene H. Dortch, Secretary  
Federal Communications Commission  
445 Twelfth Street, S.W.  
Washington, DC 20554

Re: WC Docket No. 19-195  
WC Docket No. 11-10

Madam Secretary:

In accordance with Section 1.1206(b) of the Commission's rules,<sup>1</sup> this letter provides notice of oral ex parte presentations to the Commission in the above-captioned dockets. On September 18, 2019, undersigned counsel, along with Paula Boyd, Paul Garnett, John Kahan, and Allen Kim of Microsoft Corporation, met with Allison Baker, Kirk Burgee, Joseph Calascione, Travis Litman, Steven Rosenberg, Arielle Roth, and Alisa Valentin.

In its presentations, Microsoft discussed several aspects of the Report and Order and Second Further Notice of Proposed Rulemaking adopted by the Commission in the above-captioned dockets,<sup>2</sup> as well other related issues, including Microsoft's development of a Machine Learning Model ("ML Model") that it believes can assist the Commission as well as industry and other stakeholders in detecting potential outliers that diverge from reported broadband<sup>3</sup> availability information.

Microsoft commended the Commission for its adoption of the *Report and Order* and for the issues the Commission has raised in the *Further Notice*, stating that it is encouraged by the steps the Commission has taken to improve the accuracy and granularity of broadband

---

<sup>1</sup> 47 C.F.R. § 1.1206(b).

<sup>2</sup> *Establishing the Digital Opportunity Data Collection, et al.*, WC Docket No. 19-195, *et al.*, Report and Order and Second Further Notice of Proposed Rulemaking, FCC 19-79 (rel. Aug. 6, 2019) ("*Report and Order*" or "*Further Notice*").

<sup>3</sup> Unless otherwise stated herein, "broadband" is intended to refer to broadband service providing 25 Mbps download and 3 Mbps upload speeds.

mapping data, and by the further actions it plans to pursue in its efforts to track the availability and usage of broadband services. Several topics addressed during Microsoft's meetings with Commission staff are discussed in the following paragraphs.

### **1. FCC Form 477 and Digital Opportunity Data Collection Reporting.**

Microsoft praised the decision in the *Report and Order* that, for purposes of the Digital Opportunity Data Collection ("DODC"), broadband service will be treated as actually available "if the reporting fixed service provider has a current broadband connection or it could provide such a connection within 10 business days of a customer request, without an extraordinary commitment of resources, and without construction charges or fees exceeding an ordinary service activation fee."<sup>4</sup> Microsoft observed that the 10-day criterion, and the restriction placed on construction charges and fees, are significant improvements, that will enable greater precision in determining whether broadband service is "available," and that will help to reduce overstated coverage in broadband availability reports.

Microsoft expressed concern that the Commission opted in the *Report and Order* not to make any revisions to the FCC Form 477 ("Form 477") instructions concerning reporting the availability of fixed broadband connections in a census block. As a result, the Form 477 and the reporting criteria for the DODC are now misaligned, and without any improvements to Form 477, difficulties remain in accurately determining broadband availability in the near term. Microsoft advocated that the Commission should act expeditiously to improve the accuracy of Form 477 by correcting this misalignment.

Microsoft suggested that the best way to improve the accuracy of Form 477 would be to revise Section 5.3 of the [Form 477 Instructions](#). The Instructions currently provide that a fixed broadband connection is considered "available" if the service provider "could, within a service interval that is typical for that type of connection—that is, without an extraordinary commitment of resources—provision" broadband service<sup>5</sup> Microsoft observed that this loose formulation has the disadvantage of leaving considerable discretion to the reporting service provider, and consequently has contributed to the overstatement of broadband availability. Revising the Section 5.3 formulation of availability—to mirror the new DODC criteria—would help to reduce the risk of overstated broadband coverage being reported by Form 477 filers and cure the misalignment between the Form 477 and DODC definitions.

Microsoft also advocated that the Commission should act quickly to align the Form 477 and DODC criteria for determining broadband availability, as a first step in the transition from the use of Form 477 data to the use of DODC data.

---

<sup>4</sup> *Report and Order* at para. 13 (emphasis added) (footnote omitted).

<sup>5</sup> FCC Form 477, Local Telephone Competition and Broadband Reporting, Instructions, at § 5.3.

## **2. Using a Machine Learning Model to Identify Potential Outlier ZIP Codes.**

A principal focus of Microsoft's presentations was a discussion of its development of an ML Model to predict broadband service availability, and a review of data it has generated through the use of this model. Microsoft has used its model to identify potential outlier ZIP Codes, based on comparisons with Form 477 broadband coverage data submitted to the Commission.<sup>6</sup>

In developing the ML Model, Microsoft's objective has been to design a means by which machine learning capabilities can be utilized to identify potential inaccuracies in reported broadband availability data, and thus warrant further examination by the Commission, or by industry or other stakeholders. The model is highly accurate at the national level, predicting broadband availability with 90 percent accuracy. The model's accuracy diminishes somewhat at the more granular ZIP Code level, demonstrating a capability to predict 78 percent of ZIP Codes within 10 percentage points of reported broadband availability, and a capability to predict 63 percent of ZIP Codes within 5 percentage points of reported broadband availability. These levels of accuracy at the ZIP Code level are sufficient, however, to make the ML Model a useful tool in identifying ZIP Codes that may have inaccuracies in reported broadband availability.

The ML Model is built to use usage data and demographic data to enable predictions of broadband availability, thus serving as a useful auditing and crowdsourcing tool in identifying where there may be problems with availability and broadband service quality. Microsoft utilized an extensive array of device-level data accumulated from more than 200 Microsoft services<sup>7</sup> as a basis for estimating broadband download speeds and broadband service coverage. Microsoft indicated that its usage data is extremely precise, showing the "ground truth" concerning progress being made in the adoption of broadband services across the nation.

---

<sup>6</sup> Microsoft used availability data reported to the Commission at the census tract level and "converted" this data to illustrate availability on a ZIP Code basis, using Department of Housing and Urban Development ("HUD") [USPS Zip Code Crosswalk Files](#). The HUD files were used by Microsoft to determine the percentage of each census tract that is associated with a particular ZIP Code. Microsoft chose to analyze its ML Model data on a ZIP Code basis because that is the lowest level of granularity for which Microsoft has confidence in the data generated by the model, and because consumers are more familiar with ZIP Codes than with census tract numbers.

<sup>7</sup> Microsoft did not utilize any personally identifiable information ("PII") in connection with its design and use of the ML Model.

Combined with various other demographic indicators (*e.g.*, land devoted to agriculture, education levels, median household income, home ownership),<sup>8</sup> this usage data works as a powerful predictor of broadband availability.

Microsoft noted that, currently, there are discrepancies concerning nationwide data relating to broadband availability and broadband usage. Data reported to the Commission by service providers indicates that 93.5 percent of all areas in the nation have access to fixed broadband service, while Microsoft estimates, on the basis of its usage data, that approximately 49 percent of all Americans are accessing the internet at broadband speeds. Microsoft observed that, although availability can be expected to be somewhat “north” of usage, such a substantial gap suggests issues with the availability data.

The Commission’s subscriber data and other customer survey data correlate relatively closely with Microsoft’s usage data, thus confirming Microsoft’s view that the discrepancies between broadband availability and usage data are not caused by inaccurate usage data. Microsoft indicated in its presentations that the Commission’s data shows a nationwide broadband subscribership level of 54 percent, that a Pew Research survey shows that 65 percent of respondents indicate that they are home broadband users, and that consumer responses gathered by American FactFinder (Census Bureau) show that 67 percent of consumers have broadband subscriptions.

Microsoft explained to Commission staff that it has tested the utility of its ML Model by using it to identify potential outlier ZIP Codes. Microsoft presented the top 20 ZIP Codes that its ML Model indicate are likely the most challenged with respect to broadband availability, and noted that the 20 ZIP Codes also show substantial gaps between estimated broadband availability (based on data reported to the Commission) and the estimated percentage of consumer broadband usage (based on Microsoft’s data). Microsoft also presented 12-month speed test charts (from BroadbandNow.com) for each of the 20 outlier ZIP Codes, which show that, for almost all of the months for all of the 20 ZIP Codes, available broadband speeds were below 25/3 Mbps, calling into question the extent to which broadband actually is available in any of these ZIP Codes and confirming the results of the ML Model.

Microsoft stated that the results produced by its model simply validate what we already know: Broadband availability data reported to the Commission tends to overstate availability. Microsoft expressed the view that the model, especially when combined with subscription data (see below), may serve the Commission, as well as industry and other stakeholders, as a useful auditing tool to identify where the broadband availability problems are.

---

<sup>8</sup> Among the indicators used, usage, not surprisingly, is the most important variable in predicting availability, by a considerable margin.

Microsoft indicated in its presentations that it plans to make both its ML Model, and data the model has produced, available to the public. It plans to make the model open-sourced and publicly available on GitHub, enabling all stakeholders to utilize the model, as well as giving them an opportunity to strengthen the model's performance by adding their own datasets to the model. As Microsoft continues to use the model to identify outlier ZIP Codes, it also plans to publish additional ZIP Code results after confirming those results.

### **3. Pairing Broadband Coverage Polygons/Shapefiles with Datasets.**

The *Report and Order* requires all fixed broadband providers to submit broadband coverage polygons depicting the areas where the providers actually have broadband-capable networks and make fixed broadband service available to end-user customers.<sup>9</sup> Microsoft expressed its support for this decision to require the submission of polygons, in part because the use of broadband coverage polygons will cure one of the problems, the lack of granularity, associated with current broadband availability reporting criteria.

Specifically, service providers currently report lists of census blocks in which they provide broadband service, but they are not required to report the extent of their coverage within a census block. Thus, if they serve only one location in a census block, the Commission treats the entire census block as being served. Now, broadband providers will be required to certify that broadband is available at every location within their reported polygons.

Microsoft also addressed the issue of what datasets should be employed in conjunction with the polygon maps, advocating that individual buildings are likely best as the locations that must have access to broadband. Microsoft acknowledged that the use of a buildings/structures dataset will require the resolution of various issues (e.g., whether and how to distinguish between residential and commercial buildings).

### **4. Crowdsourcing.**

Microsoft expressed its view that crowdsourced data serves as another effective tool for verifying broadband providers' service availability reports and expressed its support for the Commission's requiring the use of crowdsourcing. Microsoft suggested that the Commission should collect crowdsourced data that relates to both broadband availability and broadband usage. Microsoft noted that, based on its work with its ML Model, usage data is the most critical input into identifying where the Commission may have potential outlier areas with respect to its broadband availability dataset. Without usage data, it could be very difficult to audit the Commission's availability data effectively.

---

<sup>9</sup> *Report and Order* at para. 12.

## **5. The Use of Subscription Data.**

Microsoft expressed the view that the granular 25/3 Mbps broadband subscription data collected by the Commission should be made public, to the extent practicable. In general, subscription data can serve as an effective means of validating the accuracy of reported broadband availability. More specifically, the Commission's subscription data, if made publicly available, could be used as an additional dataset for Microsoft's ML Model, thus making the model an even more accurate predictor of availability.

\* \* \* \* \*

Copies of Microsoft's presentation slides are enclosed for the record. The "FCC1" slide deck was used in meetings with Travis Litman and Arielle Roth, the "FCC2" slide deck was used in a meeting with Allison Baker, Kirk Burgee, and Steven Rosenberg, and the "FCC3" slide deck was used in meetings with Joseph Calascione and Alisa Valentin.

Should you have any questions, please contact the undersigned directly.

Sincerely,

### **MICROSOFT CORPORATION**

Paula Boyd, Senior Director  
U.S. Government and Regulatory Affairs  
901 K Street, N.W., 11th Floor  
Washington, DC 20001  
202-263-5946

John Cimko  
Lukas, LaFuria, Gutierrez & Sachs, LLP  
8300 Greensboro Drive, Suite 1200  
Tysons, VA 22102  
703-584-8686

*Counsel to Microsoft Corporation*

Enclosures

cc: Allison Baker  
Kirk Burgee  
Joseph Calascione  
Travis Litman

Steven Rosenberg  
Arielle Roth  
Alisa Valentin  
Paul Garnett  
John Kahan  
Allen Kim